# Proyecto

***"Estudio Experimental del Efecto del Paso de Mensajes en Ambiente GRID para el Desarrollo de Sistemas que Tratan con Problemas NP-Completos"***

## REPORTE DE AVANCES DEL PRIMER SEMESTRE

### Colaboradores en el proyecto


**Universidad Autónoma del Estado de Morelos**
Centro de Investigación en Ingeniería y Ciencias Aplicadas

**Investigador Líder**
Dr. Marco Antonio Cruz Chávez
**Estudiantes**
Fredy Juárez Pérez
Erika Yesenia Avilar Melgar
Alina Martínez Oropeza
Pedro Moreno Bernal
René López Ruiz
Wendy Torres Manjarrez

**Instituto Tecnológico de Veracruz**
Departamento de Sistemas y Computación

**Investigador Principal**
Dr. Abelardo Rodríguez León
**Estudiantes**
Jonathan Damaro Lope Torres
Gabriela Pérez Reyes
Rodrigo Eugenio Morales Navarro

**Universidad Politécnica del Estado de Morelos**
Departamento de Ingeniería en Informática

**Investigador Principal**
C. Dr. Irma Yazmín Hernández Báez
**Estudiantes**
Roberto Estrada Alcazar

## 1. Objetivos

General
Diseñar, implementar y evaluar un modelo de paralelismo de tres niveles (SMP, Cluster y Grid) con paso de mensajes eficiente, aplicado a algoritmos genéticos que resuelvan problemas duros de tipo NP-Completos.

Específicos

- Definir tipos de estructuras de datos en algoritmos, adecuadas para implementar en los algoritmos desarrollados por el grupo de trabajo, ruteo de vehículos, calendarización de sistemas de manufactura.
- Diseñar e implementar un algoritmo con paralelismo a nivel de SMP y de cluster, para los tres tipos de problemas seleccionados, tomando en cuenta el tipo de estructura de datos adecuada para el paso de mensajes.
- Diseñar e implementar un algoritmo con paralelismo a nivel Grid para los tres tipos de problemas NP-Completos definidos.
- Evaluar y comparar en forma separada cada uno de los algoritmos paralelos desarrollados
- Integrar los algoritmos desarrollados en los niveles anteriores.
- Evaluar los resultados de la versión integrada.

## 2. Actividades desarrolladas

El estándar MPI define la sintaxis y la semántica de las funciones contenidas en una biblioteca de paso de mensajes diseñada para ser utilizadas en programas que exploten la existencia de múltiples procesadores. La Interfaz de Paso de Mensajes conocido ampliamente como MPI, es el estándar para la comunicación entre los nodos que ejecutan un programa en un sistema de memoria distribuida. MPI consisten en un conjunto de bibliotecas que se utilizan en programas para distintos lenguajes de programación. Las computadoras con sistemas de memoria distribuida son fáciles de escalar, mientras que la demanda de los recursos crece, se puede agregar más memoria y procesadores, esto es aprovechado por MPI para la distribución de procesos en los recursos contenidos. La desventaja principal de las MPI es el acceso remoto a memoria el cual es lento y la programación puede resultar complicada.

### 2.1. Creación de tipos de datos complejos definidos por el usuario

MPI provee flexibilidad en la construcción de nuevos tipos de datos que el usuario puede definir, como pueden ser las estructuras de datos, esta investigación esta basada en la creación de estructuras de datos complejas, es decir, estructuras que sus tipos de datos sean estructuras anidadas y que la creación de datos con estas estructuras complejas se definan utilizando memoria dinámica.
Los tipos de datos derivados, nos permiten agrupar datos de diferentes tipos y manejarlos como un sólo parámetro, además que con MPI se definen estructuras de datos heterogéneas para un mejor manejo de los recursos computacionales y comprensión del propio algoritmo

que maneja estas estructuras de datos. De acuerdo a la investigación realizada en el presente proyecto, se ha comprobado que la dependencia de datos complejos en un algoritmo no influye en la ejecución de éste. Esto último favorece el paso de mensajes y la facilidad en la programación del algoritmo. Lo que falta es realizar la medición de forma experimental de la latencia que provoca la transferencia de datos empaquetados entre clusters geográficamente distanciados con una conexión de Internet 2.
La investigación planteada permitirá conocer las características más resaltantes del rendimiento de los algoritmos con paso de mensajes en una plataforma Grid y contribuir al conocimiento de los problemas que afectan al desempeño de algoritmos e implementaciones similares, así como los puntos positivos de este enfoque.
Con los resultados de esta investigación se podrá aportar también recomendaciones, sugerencias o nuevos temas de estudio acerca del desarrollo de éste u otros algoritmos, posiblemente optimizados para plataformas de programación distribuida.

### 2.2. Migración de clusters Tarántula de Scientific Linux a Rocks Cluster e integración del cluster de la UPEMOR a la Grid Tarántula.

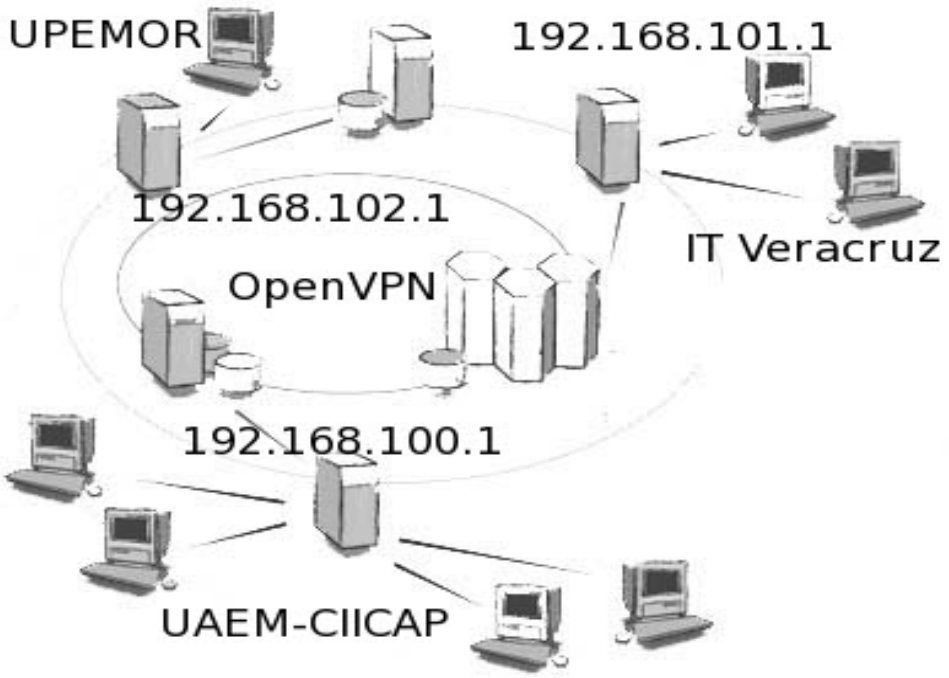
Rocks es una distribución open-source para clusters Linux de 32 y 64 bits. Permite a usuarios finales un camino fácil de construir clusters de computadoras y Grids, fácil de desarrollar, administrar, actualizar y escalar, debido a esto cientos de investigadores alrededor del mundo se encuentren utilizando rocks para desarrollar sus propios clusters.
Debido a la problemática de ejecutar jobs en plataformas con infraestructura geográficamente dispersa, el mandar a ejecutar un job distribuido donde los clusters son de 32 y 64 bits y distribuciones de Sistema Operativo diferente, se torna complejo la ejecución y obtención de los resultados y debido a este inconveniente se ve la necesidad de estandarizar una misma distribución de sistema operativo en los clusters que forman parte de la grid Tarántula, teniendo la necesidad de migrar los Clusters de la Grid Tarántula a Rocks Cluster, puesto que el IT de Veracruz cuenta con un cluster de alto rendimiento llamado Nopal, el cual tiene instalado la distribución Rolled Tacos de Rocks Cluster 5.3, y la UPEMOR esta en proceso de instalación de la misma distribución.
En la UAEM, las dificultades obtenidas con los equipos compute node, fue que las tarjetas de red no soportan booteo por PXE desde el BIOS, este problema se resolvió actualizando el bios, con una actualización al controlador de la tarjeta de red y buscando la forma desde el CD de Rocks bootear.
Hace poco tiempo la UPEMOR ha conseguido equipos Sunfire x2270W, para implementar su cluster, pero las demoras se deben a que aun no tienen asignada una dirección publica en su red privada para tener acceso y salida a Internet.
De la misma forma que la UAEM encuentra dificultades para la instalación de los compute node, el ITVeracruz paso por el mismo proceso, el cual también ya se solucionó.

### 2.3. Conexión infraestructura UAEM-ITVer-UPEMOR

Conectar a través de Open VPN los clusters de las instituciones UAEM, IT Veracruz y UPEMOR, donde a través del uso de certificados de SSH, se tiene acceso remoto y se procede a compartir recursos de procesamiento distribuido, para ejecutar jobs de un algoritmo al mismo tiempo en la Grid Tarántula.
Características técnicas Cluster Rocks UAEM.

1 equipo con procesador Pentium 4 a 2793 Mhz, 512 MB Memoria RAM, 80 GB en disco duro y 2 tarjetas de red10/100 Mbps, y 18 nodos esclavos con procesador Intel Celeron Dual Core a 2.0 Ghz, 2 GB de memoría RAM, 160 GB en disco duro y 1 tarjeta de red 10/100 Mbps.
Carcaterísticas técnicas Cluster Rocks UPEMOR.
5 equipos Sunfire x2270w con procesador Intel Core 2 Duo a 2.6 Ghz, con 2 GB en memoria RAM, disco duro de 160 GB.
Características técnicas Cluster Nopal IT Veracruz.
1 equipo con procesador pentium 4, 2394 Mhz, 512 GB Memoria RAM, 60 GB disco duro y 2 tarjetas de red 10/100 Mbps.
2 equipo con procesador Pentium 4 Dual Core, 3200 Mhz, 1 GB Memoria RAM, 80 GB disco duro y1 tarjeta de red 10/100 Mbps.

Diagrama infraestructura Grid Tarántula en conexión clusters instituciones por openVPN.




**2.4. Algoritmo paralelo multinivel del GA para la solución del problema VRPTW.**

Se revisó el algoritmo paralelo del genético mutador inteligente para VRPTW realizado en el proyecto anterior en el cual se obtuvo un algoritmo multievolutivo, ya que al ejecutar el algoritmo en un clúster cada segmento del algoritmo comienza con su propia población inicial de, 100 individuos por nodo, aumentando la zona de búsqueda, después se hace la selección el cruzamiento y la mutación. Posteriormente, se combinan individuos de cada nodo para hacer la nueva población inicial con lo cual todos los nodos obtienen la misma población de 100 individuos para la siguiente generación. Como el algoritmo con el que se está comparando es secuencial al comenzar la ejecución crea una población de 100 individuos y comienzan las iteraciones, así que se tuvo que modificar el algoritmo paralelo

para que todos los nodos empezaran con la misma población inicial que es en lo que se está trabajando.
Las herramientas AutoMap y AutoLink se utilizan para el paso de mensajes de estructuras complejas facilitando así la programación y se espera que también mejoren la eficiencia de los programas paralelos. Por el momento sólo se estudio como se instala, configura y utiliza y la forma más factible de utilizarla en el algoritmo mutador inteligente para VRPTW que se implementará. Dada la falta de experiencia al utilizar las herramientas de automap y autolink se ha tenido un retraso en la paralelización del algoritmo genético, ya que la información existente sobre automap y autolink es escasa.
Se consiguieron algunos ejemplos de automap que se han analizado para ver el funcionamiento de la herramienta. Se ha observando que el programa a paralizar necesita tener las estructuras en un archivo separado al que se le aplica automap, que genera las estructuras en tipos de datos MPI equivalentes agregándole al nombre de la estructura original el prefijo AM_ que será el que se utilice como tipo de dato en las instrucciones del paso de mensaje.
Hasta este paso el programa está en dos archivos y se necesita ejecutar el automap antes de poder ejecutar el programa paralelizado, así que se tiene que utilizar el autolink para poder enlazar las estructuras ya convertidas a tipos de datos MPI al programa paralelizado para que se pueda ejecutar. Esto se hace cambiando las sentencias de MPI a sentencias de autolink para que se pueda enviar la estructura compleja por paso de mensajes.
En el algoritmo de Vehicle Routing Problems with Time Windows (VRPTW) empleado en este proyecto, se detectaron dos métodos que son los que más tiempo de ejecución consumen. Este análisis se realizó a través de la creación de un profile. Se evaluaron cuatro métodos de ordenación para comparar los tiempos de ejecución y elegir el que realizara la ordenación en menos tiempo. Los métodos de ordenación evaluados fueron:
· Inserción binaria
· Inserción directa
· QuickSort

Cada uno se evaluó con un arreglo de 10000 elementos, luego con 100000 y, finalmente, con 1000000 para someterlos a distintas condiciones, siendo la última opción la más similar durante la ejecución del algoritmo de VRPTW por la cantidad de datos que se manejan en la ordenación. Tomando en cuenta los tiempos de ejecución para la ordenación de 1000000 de datos tipo int, los métodos menos eficientes son el de Inserción binaria e Inserción directa. El método más eficiente es el QuickSort, con un tiempo promedio de ejecución de 431 milisegundos. Por lo tanto, el QuickSort se ha elegido para ser paralelizado empleando OpenMP y, posteriormente, implementarlo en el algoritmo del VRPTW. Se realizó la paralelización del algoritmo QuickSort empleando la librería OpenMP para realizar dicha tarea, con lo cual se obtuvo una ganancia de aproximadamente 90 milisegundos, ya que el tiempo de ejecución se redujo de un promedio de 431 milisegundos a 341 milisegundos.

### 2.5. Desarrollo de los algoritmos con paralelismo a nivel de Grid

En el desarrollo del algoritmo genético para calendarización de sistemas de manufactura, utilizando un empaquetamiento de datos complejos con memoria dinámica, se trabaja de acuerdo al punto 2.1 y también se estudia las herramientas de AutoMap y AutoLink. Se espera a finales de julio tener la estructura de codificación de estos datos complejos para

implementarse en el algoritmo genético de acuerdo a los resultados obtenidos en 2.1 y con las herramientas de AutoMap.
Actualmente se está terminando la estructura de vecindad híbrida que se aplicará al genético. Las actividades realizadas para la implementación de esta estructura de vecindad son las siguientes.

| ACTIVIDAD | TRABAJO REALIZADO |
|---|---|
| EDO. DEL ARTE DE ESTRUCTURAS DE VECINDAD APLICADAS A PROBLEMAS EN TALLERES DE MANUFACTURA. | INVESTIGACIÓN SOBRE LAS ESTRUCTURAS APLICADAS AL PROBLEMA DE CALENDARIZACIÓN EN TALLERES DE MANUFACTURA, EN QUE TIPO DE ALGORITMOS SE HAN APLICADO, COMO FUNCIONAN, CUAL HA SIDO SU EFECTIVIDAD Y LA CORRESPONDIENTE EXPOSICIÓN EN SEMINARIO SOBRE LA INVESTIGACIÓN. |
| EL MÉTODO DE LA RUTA CRÍTICA O CPM (CRITICAL PATH METHOD). | ESTUDIO SOBRE COMO ENCONTRAR LA RUTA CRÍTICA EN UN GRAFO POR EL MÉTODO CPM, CÁLCULO DEL MÁXIMO TIEMPO MÁS PRÓXIMO, CÁLCULO DEL MÍNIMO TIEMPO MÁS LEJANO, CÁLCULO DE LAS HOLGURAS E IDENTIFICACIÓN DE OPERACIONES MIEMBROS DE LA RUTA CRÍTICA. |
| HIBRIDACIÓN DE UNA ESTRUCTURA DE VECINDAD CON LAS PRNCIPALES ESTRUCTURAS QUE DAN MEJORES RESULTADOS PARA EL PROBLEMA DE MANUFACTURA | APLICAR LAS ESTRUCTURAS DE VECINDAD EN LA RUTA CRITICA E IMPLEMENTARLA EN EL ALGORITMO GENÉTICO |

## 3. Ajuste de actividades, Justificación

Actualmente la UAEM esta en proceso de remodelación e instalación de la infraestructura eléctrica, debido a un cambio físico del cluster, para garantizar la seguridad y disponibilidad, el frontd end, así como 2 compute node, están disponibles en http://www.gridmorelos.uaem.mx:8080/. La UPEMOR esta en proceso de asignación de dirección IP y registro de DNS, para proceder a levantar el frontend y compute nodes, y tomando en cuenta que posiblemente se encuentren algunas dificultades a la hora de instalar Rocks sobre plataformas Sunfire de Sun Microsystems. IT Veracruz tiene completo y disponible el cluster Nopal, en estado de pruebas locales, listo para poder formar parte de la MiniGrid Tarántula que esta realizando con las instituciones antes mencionadas. Debido a lo anterior, la MiniGrid Tarántula terminara su instalación y configuración a finales de agosto para proceder a las pruebas de latencia, ancho de banda y ejecución de los algoritmos propuestos.

Dada la complejidad de la implementación de automap y autolink para el empaquetamiento de datos complejos y lo complejo del algoritmo a paralelizar se ha tenido un retraso imprevisto y se espera que en el ajuste del calendario de actividades se tenga los resultados esperados.

## 4.- Calendario de Actividades, Reprogramación y Estatus

| Actividad/Mes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estado del arte sobre la comunicación entre nodos de una Grid. **Grupo de Investigación UPEMOR** | x | x | x | | | | | | | | | |
| Estudio de la dependencia de los datos de los algoritmos genéticos utilizados para los problemas NP-Completos **Grupo de Investigación UAEM** | x | x | x | x | x | x | | | | | | |
| Integración a la Grid Tarántula del clúster de la U.P.E.Mor. **Grupo UPEMOR-UAEM-ITVer** | | | | | | | x | x | | | | |
| Revisión y propuesta de técnicas de empaquetamiento de los datos complejos para paso de mensajes con MPI. **Grupo de Investigación ITVer** | | | x | x | x | x | | | | | | |
| Implementación de los algoritmos de empaquetamiento de los datos complejos para reducir la dependencia. **Grupo de Investigación UAEM** | | | | | x | x | x | x | x | x | | |
| Desarrollo de los algoritmos con paralelismo a nivel de SMP y de cluster **Grupo UPEMOR-UAEM-ITVer** | | | | x | x | x | x | X | x | x | | |
| Desarrollo de los algoritmos con paralelismo a nivel de Grid **Grupo UPEMOR-UAEM-ITVer** | | | | x | X | x | x | x | x | | | |
| Pruebas experimentales a nivel SMP y de cluster **Grupo UPEMOR-UAEM-ITVer** | | | | x | x | x | x | x | x | x | x | |
| Pruebas experimentales en la Grid Tarántula compuesta de tres clúster **Grupo UPEMOR-UAEM-ITVer** | | | | | | | x | x | x | x | x | |
| Desarrollo del reporte de dependencias y de eficiencia de los algoritmos ejecutados en los tres niveles **Grupo UPEMOR-UAEM-ITVer** | | | | | | | | | | x | x | x |
| Elaboración de artículos relacionados al proyecto **Grupo UPEMOR-UAEM-ITVer** | | | | x | x | x | | | | x | x | x |